

RAW SEQUENCE LISTING

**The Biotechnology Systems Branch of the Scientific and Technical
Information Center (STIC) no errors detected.**

Application Serial Number: 09/936,823A
Source: 1FW16
Date Processed by STIC: 10/12/05

ENTERED



IFW16

RAW SEQUENCE LISTING

DATE: 10/12/2005

PATENT APPLICATION: US/09/936,823A

TIME: 11:23:19

Input Set : A:\2005-10-06 0933-0170P.ST25.TXT

Output Set: N:\CRF4\10122005\I936823A.raw

3 <110> APPLICANT: Merja PENTTILA et al.
 5 <120> TITLE OF INVENTION: PROCESS FOR PARTITIONING OF PROTEINS
 7 <130> FILE REFERENCE: 0933-0170P
 9 <140> CURRENT APPLICATION NUMBER: US 09/936,823A
 10 <141> CURRENT FILING DATE: 2001-10-24
 12 <150> PRIOR APPLICATION NUMBER: PCT/FI00/00249
 13 <151> PRIOR FILING DATE: 2000-03-24
 15 <150> PRIOR APPLICATION NUMBER: FI 19991782
 16 <151> PRIOR FILING DATE: 1999-08-20
 18 <150> PRIOR APPLICATION NUMBER: FI 19990667
 19 <151> PRIOR FILING DATE: 1999-03-25
 21 <160> NUMBER OF SEQ ID NOS: 46
 23 <170> SOFTWARE: PatentIn Ver. 2.2
 25 <210> SEQ ID NO: 1
 26 <211> LENGTH: 428
 27 <212> TYPE: DNA
 28 <213> ORGANISM: Trichoderma reesei
 30 <220> FEATURE:
 31 <221> NAME/KEY: intron
 32 <222> LOCATION: (167)..(236)
 34 <220> FEATURE:
 35 <221> NAME/KEY: intron
 36 <222> LOCATION: (323)..(386)
 38 <220> FEATURE:
 39 <221> NAME/KEY: misc_feature
 40 <223> OTHER INFORMATION: Coding sequence of hfb1
 42 <400> SEQUENCE: 1
 43 atgaagttct tcgccatcgc cgctctcttt gccgcgctg ccgttgccca gcctctcgag 60
 44 gaccgcagca acggcaacgg caatgtttgc cctcccgcc tcttcagcaa cccccagtgc 120
 45 tgtgccaccc aagtccttgg cctcatcggc cttgactgca aagtcgtaa gttgagccat 180
 46 aacataagaa tcctcttgac ggaaatatgc cttctcactc ctttaccct gaacagcctc 240
 47 ccagaacggt tacgacggca ccgacttccg caacgtctgc gccaaaaccg gcgcccagcc 300
 48 tctctgctgc gtggcccccg ttgtaagttg atgccccagc tcaagctcca gtctttggca 360
 49 aaccattct gacaccaga ctgcaggccg gccaggctct tctgtgccag accgccgtcg 420
 50 gtgcttga 428
 53 <210> SEQ ID NO: 2
 54 <211> LENGTH: 78
 55 <212> TYPE: DNA
 56 <213> ORGANISM: Artificial Sequence
 58 <220> FEATURE:
 59 <223> OTHER INFORMATION: Description of Artificial Sequence: PCR 5' primer
 61 <400> SEQUENCE: 2
 62 tcgggcacta cgtgccagta tagcaacgac tactactcgc aatgccttgt tccgcgtggc 60

P-6

RAW SEQUENCE LISTING

DATE: 10/12/2005

PATENT APPLICATION: US/09/936,823A

TIME: 11:23:19

Input Set : A:\2005-10-06 0933-0170P.ST25.TXT

Output Set: N:\CRF4\10122005\I936823A.raw

```

63      tctagttctg gaaccgca                                78
66 <210> SEQ ID NO: 3
67 <211> LENGTH: 30
68 <212> TYPE: DNA
69 <213> ORGANISM: Artificial Sequence
71 <220> FEATURE:
72 <223> OTHER INFORMATION: Description of Artificial Sequence: PCR 3' primer
74 <400> SEQUENCE: 3
75      tcgtacggat cctcaagcac cgacggcggg                    30
78 <210> SEQ ID NO: 4
79 <211> LENGTH: 63
80 <212> TYPE: DNA
81 <213> ORGANISM: Artificial Sequence
83 <220> FEATURE:
84 <223> OTHER INFORMATION: Description of Artificial Sequence: PCR 5' primer
86 <400> SEQUENCE: 4
87      actacacgga ggagctcgac gacttcgagc agcccagct gcacgcagag caacggcaac 60
88      ggc                                                    63
91 <210> SEQ ID NO: 5
92 <211> LENGTH: 2211
93 <212> TYPE: DNA
94 <213> ORGANISM: Trichoderma reesei
96 <220> FEATURE:
97 <221> NAME/KEY: promoter
98 <222> LOCATION: (1)..(2211)
99 <223> OTHER INFORMATION: cbh1 promoter sequence
101 <400> SEQUENCE: 5
102      gaattctcac ggtgaatgta ggccttttgt agggtaggaa ttgtcactca agcaccceca 60
103      acctccatta cgcctcccc atagagttcc caatcagtga gtcattggac tgttctcaaa 120
104      tagattgggg agaagttgac ttccgcccag agctgaaggc cgcacaaccg catgatatag 180
105      ggctcggaac ggcaaaaaag cacgtggctc accgaaaagc aagatgtttg cgatctaaca 240
106      tccaggaacc tggatacatc catcatcacg caccgacct ttgatctgct ggtaaactcg 300
107      tattcgccct aaaccgaagt gcgtggtaaa tctacacgtg ggccccttct ggtatactgc 360
108      gtgtgtcttc tctaggtgca ttctttcctt cctctagtgt tgaattgttt gtgttgggag 420
109      tccgagctgt aactacctct gaatctctgg agaattggtg actaacgact accgtgcacc 480
110      tgcattcatgt atataatagt gatcctgaga agggggggtt ggagcaatgt gggactttga 540
111      tggatcatca acaaagaacg aagacgcctc ttttgcaaag ttttgtttcg gctacgggtga 600
112      agaactggat acttgttgtg tcttctgtgt atttttgtgg caacaagagg ccagagacaa 660
113      tctattcaaa caccaagctt gctcttttga gctacaagaa cctgtggggg atatattctag 720
114      agttgtgaag tcggtaatcc cgctgtatag taatacagag cgcattctaa tactccgaag 780
115      ctgctgcgaa cccggagaat cgagatgtgc tggaaagctt ctacgcagcg gctaaattag 840
116      catgaaaggc tatgagaaat tctggagacg gcttgtttga tcatggcggt ccattcttcg 900
117      acaagcaaag cgttccgctg cagtagcagg cactcattcc cgaaaaaact cggagattcc 960
118      taagtagcga tggaaaccga ataataaat aggcaataca ttgagttgcc tcgacgggtg 1020
119      caatgcaggg gtactgagct tggacataac tgttccgtac cccacctctt ctcaaccttt 1080
120      ggcgtttccc tgattcagcg taccggtaca agtcgtaatc actattaacc cagactgacc 1140
121      ggacgtgttt tgcccttcat ttggagaaat aatgtcattg cgatgtgtaa tttgcctgct 1200
122      tgaccgactg gggctgttcg aagcccgaat gtaggattgt tatccgaact ctgctcgtag 1260
123      aggcattgtt tgaatctgtg tcgggcagga caccgctcga aggttcacgg caagggaaac 1320

```

RAW SEQUENCE LISTING

DATE: 10/12/2005

PATENT APPLICATION: US/09/936,823A

TIME: 11:23:19

Input Set : A:\2005-10-06 0933-0170P.ST25.TXT

Output Set: N:\CRF4\10122005\I936823A.raw

```

124 caccgatagc agtgtctagt agcaacctgt aaagccgcaa tgcagcatca ctggaaaata 1380
125 caaaccaatg gctaaaagta cataagttaa tgcctaaaga agtcatatac cagcgggctaa 1440
126 taattgtaca atcaagtggc taaacgtacc gtaatttgcc aacgcgttgt ggggttgagc 1500
127 aagcaacggc aaagccact tcccacgttt gtttcttcac tcagtcgaat ctgagctggt 1560
128 gatcccccaa ttgggtcgct tgtttgttcc ggtgaagtga aagaagacag aggtaagaat 1620
129 gtctgactcg gagcggtttg catacaacca agggcagtga tgggaagacag tgaaatggtg 1680
130 acattcaagg agtatattagc cagggatgct tgagtgtatc gtgtaaggag gtttgtctgc 1740
131 cgatacgacg aatactgtat agtcacttct gatgaagtgg tccatattga aatgtaagtc 1800
132 ggcactgaac aggcataaga ttgagttgaa actgcctaag atctcgggcc ctcgggcttc 1860
133 ggctttgggt gtacatgttt gtgctcggg caaatgcaa gtgtggtagg atcgacacac 1920
134 tgctgccttt accaagcagc tgagggtatg tgataggcaa atgttcaggg gccactgcat 1980
135 ggtttcgaat agaaagagaa gcttagccaa gaacaatagc cgataaagat agcctcatta 2040
136 aacgaaatga gctagtaggc aaagtcagcg aatgtgtata tataaagggt cgagggtcgt 2100
137 gcctccctca tgctctcccc atctactcat caactcagat cctccaggag acttgtagac 2160
138 catcttttga ggcacagaaa cccaatagtc aaccgcgagc tgcgcacatc g 2211
141 <210> SEQ ID NO: 6
142 <211> LENGTH: 1588
143 <212> TYPE: DNA
144 <213> ORGANISM: Trichoderma reesei
146 <220> FEATURE:
147 <221> NAME/KEY: misc_feature
148 <223> OTHER INFORMATION: T. reesei eg11 cDNA
150 <400> SEQUENCE: 6
151 cccccctatc ttagtccttc ttgttgtccc aaaatggcgc cctcagttac actgcccgttg 60
152 accacggcca tcctggccat tgcccggctc gtgcgcgccc agcaaccggg taccagcacc 120
153 cccgaggtcc atcccaagtt gacaacctac aagtgtacaa agtccggggg gtgcgtggcc 180
154 caggacacct cgggtggtcct tgactggaac taccgctgga tgcacgacgc aaactacaac 240
155 tcgtgcaccg tcaacggcgg cgtcaacacc acgctctgcc ctgacgaggc gacctgtggc 300
156 aagaactgct tcatcgaggg cgtcgactac gccgcctcgg gcgtcacgac ctcgggcagc 360
157 agcctcacca tgaaccagta catgcccagc agctctggcg gctacagcag cgtctctcct 420
158 cggctgtatc tcctggactc tgacgggtgag tacgtgatgc tgaagctcaa cggccaggag 480
159 ctgagcttcg acgtcgacct ctctgctctg ccgtgtggag agaacggctc gctctacctg 540
160 tctcagatgg acgagaacgg gggcgccaac cagtataaca cggccggtgc caactacggg 600
161 agcggctact gcgatgctca gtgcccgcgc cagacatgga ggaacggcac cctcaacact 660
162 agccaccagg gcttctgctg caacgagatg gatatcctgg agggcaactc gagggcgaat 720
163 gccttgaccc ctcaactctg cacggccacg gccctgcgact ctgccggttg cggcttcaac 780
164 ccctatggca cgggctacaa aagctactac ggcccggag ataccgttga cacctcaag 840
165 accttcacca tcatcaccca gttcaacacg gacaacggct cgccctcggg caaccttgtg 900
166 agcatcaccc gcaagtacca gcaaacggc gtgcacatcc ccagcgccca gccggcggc 960
167 gacaccatct cgtcctgccc gtccgcctca gcctacggcg gcctcgccac catgggcaag 1020
168 gccctgagca gcggcatggt gctcgtgttc agcatttgga acgacaacag ccagtacatg 1080
169 aactggctcg acagcggcaa cgccggcccc tgcagcagca ccgagggcaa cccatccaac 1140
170 atcctggcca acaaccccaa cagcacgctc gtcttctcca acatccgctg gggagacatt 1200
171 ggggtctact cgaactcgac tgcgcccccg ccccgccctg cgtccagcac gacgttttcg 1260
172 actacacgga ggagctcgac gacttcgagc agcccagct gcacgcagac tactggggg 1320
173 cagtgcggtg gcattgggta cagcgggtgc aagacgtgca cgtcgggcac tacgtgccag 1380
174 tatagcaacg actactactc gcaatgcctt tagagcgttg acttgccctt ggtctgtcca 1440
175 gacgggggca cgatagaatg cgggcacgca gggagctcgt agacattggg cttaatatat 1500
176 aagacatgct atgttgtatc tacattagca aatgacaaac aaatgaaaaa gaacttatca 1560

```

RAW SEQUENCE LISTING

DATE: 10/12/2005

PATENT APPLICATION: US/09/936,823A

TIME: 11:23:19

Input Set : A:\2005-10-06 0933-0170P.ST25.TXT

Output Set: N:\CRF4\10122005\I936823A.raw

```

177      agcaaaaaaaaa aaaaaaaaaa aaaaaaaaa      1588
180 <210> SEQ ID NO: 7
181 <211> LENGTH: 745
182 <212> TYPE: DNA
183 <213> ORGANISM: Trichoderma reesei
185 <220> FEATURE:
186 <221> NAME/KEY: terminator
187 <222> LOCATION: (1)..(745)
188 <223> OTHER INFORMATION: T. reesei cbhl terminator
190 <400> SEQUENCE: 7
191      ggacctaccc agtctcacta cggccagtgc ggcggtattg gctacagcgg cccacaggtc 60
192      tgcgccagcg gcacaacttg ccaggctcctg aacccttact actctcagtg cctgtaaagc 120
193      tccgtgcgaa agcctgacgc accggtagat tcttggtgag cccgtatcat gacggcgggcg 180
194      ggagctacat ggccccgggt gatttatttt ttttgatatct acttctgacc cttttcaaat 240
195      atacgggtcaa ctcatctttc actggagatg cggcctgctt ggtattgcga tgttgtcagc 300
196      ttggcaaatt gtggcttttcg aaaacacaaa acgattcctt agtagccatg cattttaaga 360
197      taacggaata gaagaaagag gaaattaaaa aaaaaaaaaa aacaaacatc ccgttcataa 420
198      cccgtagaat cgccgctctt cgtgtatccc agtaccacgt caaagggtatt catgatcggt 480
199      caatgttgat attgttcgcg cagtatggct ccacccccat ctccgcgaat ctctcttctt 540
200      cgaacgcggt agtggctgct gccaatgggt aatgaccata gggagacaaa cagcataata 600
201      gcaacagtgg aaattagtgg cgcaataatt gagaacacag tgagaccata gctggcgggc 660
202      tggaaagcac tggtggagac caacttgtcc gttgcgaggc caacttgcac tgctgtcaag 720
203      acgatgacaa cgtagccgag gacccc      745
206 <210> SEQ ID NO: 8
207 <211> LENGTH: 10
208 <212> TYPE: DNA
209 <213> ORGANISM: Artificial Sequence
211 <220> FEATURE:
212 <223> OTHER INFORMATION: Description of Artificial Sequence: annealed primer
214 <400> SEQUENCE: 8
215      taaccgcggt      10
218 <210> SEQ ID NO: 9
219 <211> LENGTH: 16
220 <212> TYPE: DNA
221 <213> ORGANISM: Artificial Sequence
223 <220> FEATURE:
224 <223> OTHER INFORMATION: Description of Artificial Sequence: annealed primer
226 <400> SEQUENCE: 9
227      ctagaccgcg gttaat      16
230 <210> SEQ ID NO: 10
231 <211> LENGTH: 1232
232 <212> TYPE: DNA
233 <213> ORGANISM: Trichoderma reesei
235 <220> FEATURE:
236 <221> NAME/KEY: promoter
237 <222> LOCATION: (1)..(1232)
238 <223> OTHER INFORMATION: T. reesei gpdl promotor
240 <400> SEQUENCE: 10
241      gtcgacacga tatacaggcg cggctgatga taatgatgat cgagcatgac ttgatgctgt 60

```

RAW SEQUENCE LISTING

DATE: 10/12/2005

PATENT APPLICATION: US/09/936,823A

TIME: 11:23:19

Input Set : A:\2005-10-06 0933-0170P.ST25.TXT

Output Set: N:\CRF4\10122005\I936823A.raw

```

242  atgtgacaat attgactgcg aggaaccatc aggtgtgtat ggatggaatc attctgtaac 120
243  caccaagggtg catgcatcat aaggattctc ctcagctcac caacaacgaa cgatggccat 180
244  gttagtgaag gcaccgtgat ggcaagatag aaccactatt gcactctgcgc tccccacgca 240
245  cagtacgtca agtaacgtca aagccgacct cccgtaacct cgcccgttgt tgtccccccc 300
246  gattgcctca atcacatagt acctacctat gcattatggg cggcctcaac ccaccccccc 360
247  agattgagag ctaccttaca tcaatatggc cagcacctct tcggcgatac atactcgcca 420
248  ccccgaccgg cgcgattgtg tgtactagggt aggtcgttac tataccagca ggagagggtgc 480
249  tgcttggcaa tcgtgctcag ctgttaggtt gtacttgtat ggtacttgta aggtgggtcat 540
250  gcagttgcta aggtacctag ggagggattc aacgagccct gcttccaatg tccatctgga 600
251  taggatggcg gctggcgggg ccgaagctgg gaactcgcca acagtcatat gtaatagctc 660
252  aagttgatga taccgttttg ccagattaga tgcgagaagc agcatgaatg tcgctcatcc 720
253  gatgccgcat caccgttgtg tcagaaacga ccaagctaag caactaagggt accttaccgt 780
254  ccactatctc aggtaaccag gtactaccag ctaccctacc tgccgtgcct acctgcttta 840
255  gtgttaatat ttcacacctc ctctcaate ttcttttccc tctctctctc tttttttttt 900
256  ctctctctct ttcttctcca taaccattcc taacaacatc gacattctct cctaataacc 960
257  agcctcgcaa atcctcagtt tgtatgtacg tacgtactac aatcatcacc acgatcgctc 1020
258  gcccgacgat gcggttctg ttgcctgccc cctctctctc ctctgcccct tgacgagcta 1080
259  gcccgcgcag gactctctct cgtcaccaat ttttttccct atttaccctt cctccctctc 1140
260  tcctctctgt ttcttcttaa caaacaacca ccaccaaatt ctctttggaa gctcacgact 1200
261  cacgcagctc aattcgaga tacaatatga ga 1232

```

264 <210> SEQ ID NO: 11

265 <211> LENGTH: 1129

266 <212> TYPE: DNA

267 <213> ORGANISM: Trichoderma reesei

269 <220> FEATURE:

270 <221> NAME/KEY: terminator

271 <222> LOCATION: (1)..(1129)

272 <223> OTHER INFORMATION: T. reesei gpd1 terminator

274 <400> SEQUENCE: 11

```

275  ggatccccgag cattgtctat gaatgcaaac aaaaatagta aataaatagt aattctggcc 60
276  atgacgaata gagccaatct gctccacttg actatcttgt gactgtatcg tatgtcgaa 120
277  ccttgactgc ccattcaaac aattgtaaag gaatatagct acaagttatg tctcacgttt 180
278  gcgtgcgagc ccgtttgtac gttattttga gaaagcggtg ccatcacatg ctcacagtca 240
279  cttggcttac gatcatgttt gcgatcttcg gtaagaatac acagagtaac gattatctcc 300
280  atcgcttcta tgattaggta ctcagacaac acatgggaaa caagataacc atcgcatgca 360
281  aggtcgattc caatcatgat ctggactggg gtattccatc taagccatag taccctcgag 420
282  agaaggaatg gtaggacctc tcaggcgctc accatctgtg ctgcaaatcc aagaaacccc 480
283  ccaaaagcac ctacctatct acctagagta actgcacgag aaaagaaaag gagcagaaga 540
284  agaatgatct caagaggccg tgaacgcaga aacacactcc tcccaacttt tcaagttttg 600
285  aacaaaaaaa gaaagatgag gactagaaga tggagtattt ccttcttaga gagctctcgg 660
286  tgaggtgacc tgtcagggtt taccgcaaac cgtcggttgt tctatccaat taatcaagtc 720
287  ccgcgcctcg cctcttctct cctgtccttt catagaatcc cgtctccttg ttgcttgatc 780
288  gaagcggggt tatcgacgcc accaaagatc ttgtcttggt gacttatcaa tcctttgggtg 840
289  atcaaacagc ccccgagtga tcagatccgt aaaagaagaa gaagagtacg atttaaccag 900
290  accgaggaac aataaagcga gtaaataaca tcaaaataag agtctcgttg aaaattactt 960
291  gttcctcaat caatcccaac cccctaaaa gcccttcccc ccatgggtata tcccggcagt 1020
292  aggagagaga tatttccact accgctcacc accaagttag gcttgccgag agaagaggat 1080
293  gaatcagaag tgacaacaac gggtttagca catgggatat cggcgcgcc 1129

```

296 <210> SEQ ID NO: 12

RAW SEQUENCE LISTING ERROR SUMMARY
PATENT APPLICATION: US/09/936,823A

DATE: 10/12/2005
TIME: 11:23:20

Input Set : A:\2005-10-06 0933-0170P.ST25.TXT
Output Set: N:\CRF4\10122005\I936823A.raw

Please Note:

Use of n and/or Xaa have been detected in the Sequence Listing. Please review the Sequence Listing to ensure that a corresponding explanation is presented in the <220> to <223> fields of each sequence which presents at least one n or Xaa.

Seq#:12; N Pos. 2570,2778,2869

Invalid Line Length:

The rules require that a line not exceed 72 characters in length. This includes spaces.

Seq#:1; Line(s) 1

Seq#:42; Line(s) 1095

VERIFICATION SUMMARY

DATE: 10/12/2005

PATENT APPLICATION: US/09/936,823A

TIME: 11:23:20

Input Set : A:\2005-10-06 0933-0170P.ST25.TXT

Output Set: N:\CRF4\10122005\I936823A.raw

L:368 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:12 after pos.:2520

M:341 Repeated in SeqNo=12